



icbi

THE INNOVATION CENTER FOR BIOMEDICAL INFORMATICS

## **G-DOC *Plus* Help**

# **How to perform gene expression based survival analysis**

*Innovation Center for Biomedical Informatics  
Georgetown University*

# Overview

- Login
- Navigation
- How to create patient groups
- Different ways of performing survival analysis on gene expression data (*name of tool: Gene expression KM plot*)
  - Going from Group Comparison to Gene expression KM
  - Create groups, perform Gene expression KM
  - Going to Gene expression KM directly – run on all samples

# Login



The Innovation Center for Biomedical Informatics (ICBI)  
Lombardi Comprehensive Cancer Center

Thu Jan 22, 2015

kb472

.....

[Log In](#)

[register now](#) | [forgot password](#)

## Welcome to GDOC Plus Beta!

The Georgetown Database of Cancer Plus other diseases (G-DOC Plus) is a precision medicine platform containing molecular and clinical data from thousands of patients and cell lines, along with tools for analysis and data visualization. The platform enables the integrative analysis of multiple data types to understand disease

## Precision Medicine

## Translational research

## Population genetics

## Understanding Data in G-DOC Plus

## It all begins with a study...

All data in G-DOC Plus derives from studies on topics such as breast cancer, wound healing, or even 1,000 Genomes. Each study may contain clinical and/or biospecimen data. Below is an overview of studies by topic.

\* private studies, ones which are uploaded and marked private, are not counted here

## News

October 02, 2014: ICBI Symposium 2014

[read]

May 02, 2014: Featured in Frontiers' Top 10  
2013 Most viewed Genetics Research articles

[read]

March 12, 2014: AAAS Big Data Blog [\[read\]](#)

# G-DOC Plus Launch Pad!

Welcome! The G-DOC Plus Launch Pad is your one-stop resource for learning more about G-DOC and getting started on the platform.



Studies



Lists



**It All Starts Here!**



G-DOC has over seventy studies, We know this can be overwhelming! Let us guide you to choose the study that is relevant for your research.

[Let's Go! >](#)



Groups



Notifications **0**

# What's your area of interest?

G-DOC Plus has three overlapping entry points for the user based on their interests. Choose your area of interest to launch the workflow.



## Precision Medicine

Patients' molecular diagnostics and clinical data.



## Translational Research

Analytic tools and workflows to enable discovery.



## Population Genetics

Race-based, genomic reporting and comparison.



# Select disease/data of interest

## DEMENTIA

**1**  
*study*

**30**  
*samples*

**30**  
*biospecimen*

## PEDIATRIC CANCERS

**5**  
*studies*

**256**  
*samples*

**256**  
*biospecimen*

## BREAST CANCER

**25**  
*studies*

**3952**  
*samples*

**4532**  
*biospecimen*

## COLON CANCER

**10**  
*studies*

**1226**  
*samples*

**1262**  
*biospecimen*

## LUNG CANCER

**1**  
*study*

**478**  
*samples*

**443**  
*biospecimen*

## MUSCULAR DYSTROPHY

**1**  
*study*

**36**  
*samples*

**36**  
*biospecimen*

## LIVER CANCER

**3**  
*studies*

**298**  
*samples*

**468**  
*biospecimen*

## CELL\_LINE\_COLLECTIONS

**1**  
*study*

**60**  
*samples*

**59**  
*biospecimen*

## OVARIAN CANCER

**1**  
*study*

**1711**  
*samples*

**564**  
*biospecimen*

## PROSTATE CANCER

**1**  
*study*

**465**  
*samples*

**538**  
*biospecimen*

## PANCREATIC CANCER

**1**  
*study*

**52**  
*samples*

**51**  
*biospecimen*

## BRAIN CANCER

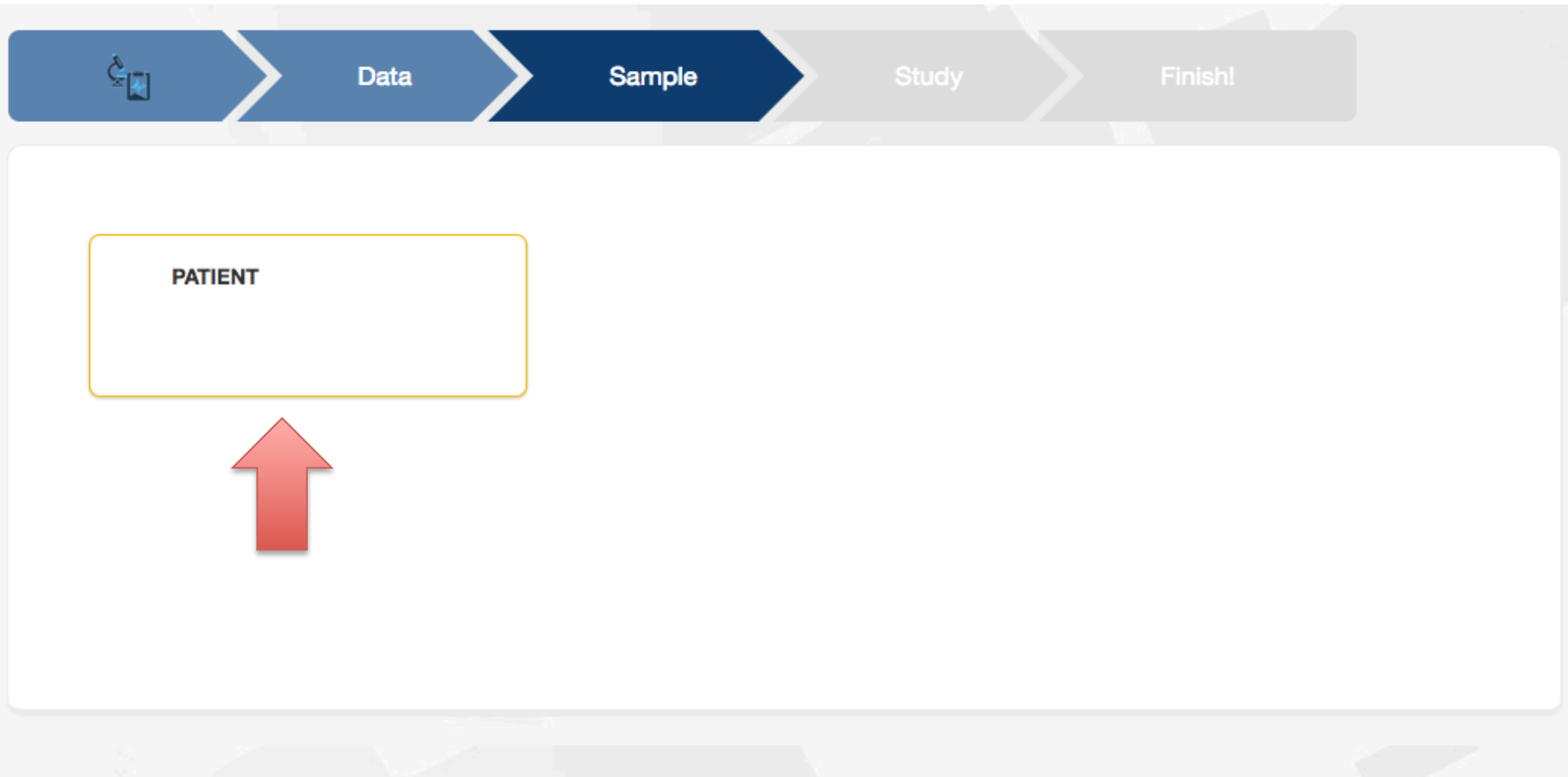
**1**  
*study*

**671**  
*samples*

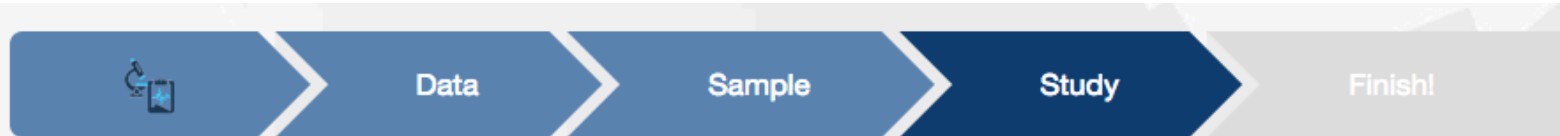
**804**  
*biospecimen*



# Choose between patient and cell line data



# Select study



## REMBRANDT

**Title:** NCI Rembrandt Study: Molecular Analysis of Brain Neoplasia

**Data Type Details:**  
CLINIC,MICROARRAY,COPY\_NUMBER

**Abstract:** This is the NCI Rembrandt Study: Molecular Analysis of Brain Neoplasia. Primary brain tumors are the fourth leading cause of cancer mortality in adults under the age of 54 years and the leading cause of cancer mortality in children in the United

**671**  
*samples*

**804**  
*biospecimen*

[More>>](#)



Note: If you click on “More”, you will see complete description of the dataset



# Study selected. Let's now create some groups

[Data](#)[Sample](#)[Study](#)[Finish!](#)

Based upon the study you picked, here is a list of tools you can use:

## Analyze

- [Classification](#)
- [KM Clinical Plot](#)
- [HeatMap Viewer](#)
- [KM Gene Expression Plots](#)
- [Group Comparison](#)
- [Chromosomal Instability Index](#)

## Search

- [Gene Expression Data](#)
- [Explore Clinical Data and Create Groups](#)

Explore clinical data and create groups



# Explore clinical data or create groups

G-DOC<sup>plus</sup>

Home Studies Lists Analyses Groups Notifications Study Options ▾ Help

Filter [reset | tips | advanced search]

**Demographics**

- ☐ Age range ⓘ
- ☐ Gender ⓘ
- ☐ Race ⓘ

**Sample details**

- ☐ Anti convulsant status ⓘ
- ☐ Copy number data ⓘ
- ☐ Gene expression data ⓘ

view all (4 more ...)

- ☐ Steroid dose status ⓘ
- ☐ Institution name ⓘ
- ☐ Prior therapy surgery done at first follow up ⓘ
- ☐ Prior therapy surgery done at second follow up ⓘ

**Disease details**

- ☒ Disease ⓘ

**Subject Search**

WHO\_GRADE(All) > DISEASE\_TYPE(All) >

Current Split Attribute ⓘ

WHO grade

Disease	WHO_GRADE_I	WHO_GRADE_II	WHO_GRADE_III	WHO_GRADE_IV
OLIGODENDROGLIOMA	0	40	25	0
ASTROCYTOMA	0	66	63	0
UNKNOWN	0	0	1	0
UNCLASSIFIED	0	0	0	0
NON TUMOR	0	0	0	0
GBM	2	0	0	140
MIXED	0	4	4	0
Total	2	110	93	140

and uncheck the boxes on the left – this will update the patient numbers in the table.

e.g. In this example you can see that among Oligodendroglioma patients, 40 patients have WHO tumor grade II and 25 patients have grade III tumor

# Let's save the Oligodendroglioma patients with tumor grade II and grade III

Filter [reset | tips | advanced search]

## Demographics

- ☐ Age range
- ☐ Gender
- ☐ Race

## Sample details

- ☐ Anti convulsant status
- ☐ Copy number data
- ☐ Gene expression data
- view all (4 more ...)
- ☐ Steroid dose status
- ☐ Institution name
- ☐ Prior therapy surgery done at first follow up
- ☐ Prior therapy surgery done at second follow up

## Disease details

- ☒ Disease

## Clinical evaluation

- ☐ Disease evaluation by MRI
- ☐ Neurologic exam score
- ☐ Performance Status Score: Karnofsky

## Subject Search

WHO\_GRADE(All) > DISEASE\_TYPE(All) >

Current Split Attribute

WHO grade

Disease	WHO_GRADE_I	WHO_GRADE_II	WHO_GRADE_III	WHO_GRADE_IV
OLIGODENDROGLIOMA	0	<a href="#">40</a>	<a href="#">25</a>	0
ASTROCYTOMA	0	<a href="#">66</a>	0	0
UNKNOWN	0	0	0	0
UNCLASSIFIED	0	0	0	0
NON TUMOR	0	0	0	0
GBM	<a href="#">2</a>	0	0	<a href="#">140</a>
MIXED	0	<a href="#">4</a>	<a href="#">4</a>	0
Total	<a href="#">2</a>	<a href="#">110</a>	<a href="#">93</a>	<a href="#">140</a>

View Detailed Report

Save ids as list

To save a list of patients, click on the hyperlinked number, and select "Save ids as list".

Repeat this procedure to create as many number of patient groups as needed.

Note: "View detailed report" shows a detailed clinical report of those selected patients.

# Enter a name for the list, and click “Save”

## Subject Search

Filter [reset] [tips] [advanced search]

### Demographics

☐ Age range ⓘ

☐ Gender ⓘ

☐ Race ⓘ

### Sample details

☐ Anti convulsant status ⓘ

☐ Copy number data ⓘ

☐ Gene expression data ⓘ

view all (4 more ...)

☐ Steroid dose status ⓘ

☐ Institution name ⓘ

☐ Prior therapy surgery done at first follow up ⓘ

☐ Prior therapy surgery done at second follow up ⓘ

### Disease details

☒ Disease ⓘ

### Clinical evaluation

☐ Disease evaluation by MRI ⓘ

☐ Neurologic exam score ⓘ

☐ Performance Status Score: Karnofsky ⓘ

WHO\_GRADE(All) > DISEASE\_TYPE(All) >

Current Split Attribute ⓘ

WHO grade ▾

Disease	WHO_GRADE_I	WHO_GRADE_II	WHO_GRADE_III	WHO_GRADE_IV
OLIGODENDROGLIOMA	0	40	25	0
			63	0
			1	0
			0	0
			0	0
			0	140
			4	0
			93	140

### Save your list

List Type:	clinical, patient
List Name:	<input type="text" value="Rem_olig_grade2"/>
<input type="button" value="Cancel"/> <input type="button" value="Save"/> <a href="#">close</a>	

Perform Group Comparison, then perform Gene expression KM analysis

## **METHOD 1**

# Now we want to compare the Oligodendroglioma tumor grade II patients with Oligo grade III patients

**G-DOC<sup>Plus</sup>** Home Studies Lists Analyses Groups Notifications Study Options ▾ Help

Filter [ reset | tips | advanced search ]

**Demographics**

- ☐ Age range
- ☐ Gender
- ☐ Race

**Sample details**

- ☐ Anti convulsant status
- ☐ Copy number data
- ☐ Gene expression data

view all (4 more ...)

- ☐ Steroid dose status
- ☐ Institution name
- ☐ Prior therapy surgery done at first follow up
- ☐ Prior therapy surgery done at second follow up

**Disease details**

- ☒ Disease

**Clinical evaluation**

- ☐ Disease evaluation by MRI
- ☐ Neurologic exam score
- ☐ Performance Status Score: Karnofsky

**Subject Search**

WHO\_GRADE(All) > DISEASE\_TYPE

Current Split Attribute

WHO grade

Disease	WHO
OLIGODENDROGLIOMA	0
ASTROCYTOMA	0
UNKNOWN	0
UNCLASSIFIED	0
NON TUMOR	0
GBM	2
MIXED	0
Total	2

Study Selected

REMBRANDT

SEARCH

- Genome Browser
- Compound/Drug Targets
- Findings
- Explore Clinical Data and Create Groups
- Gene Expression Data

ANALYZE

- Group Comparison**
- Chromosomal Instability Index
- KM Clinical Plot
- KM Gene Expression Plots
- Classification
- HeatMap Viewer

GRADE_III	WHO_GRADE_IV
0	0
0	0
0	0
0	0
140	0
0	140

# Perform Group Comparison Analysis

Current Study: REMBRANDT [change study?](#)

Select a baseline group and a comparison group(s)

Select baseline group

Rem\_olig\_grade2 ▾

p-value

.05

Multiple Comparison Adjustment

False Discovery Rate(FDR): B ▾

Select comparison group

Rem\_olig\_grade3 ▾

Fold Change

1.5

Data-Type

GENE EXPRESSION ▾

Statistical Method

T-Test: Two Sample Test ▾

Dataset

mas5 normalization ▾

Submit Analysis

Select baseline group (less screwed up group), comparison group, and settings for the comparison analysis.

Click “Submit analysis”

# Notifications

Below are your latest running analyses. Once completed, click on the Analysis name to see detailed results.

GROUP\_COMPARISON (2:05 1/22/2015 )

Complete



Once you click “submit analysis”, you will be re-directed to the “Notifications” page. Once the status is “Complete”, click on “Group comparison”



# Results of group comparison

## Analysis Results

Current Study: REMBRANDT

Analysis Results	
Statistical Method	TTest
Adjustment	FDR
Fold Change	1.5
Pvalue	.05
Study	REMBRANDT
Data File	REMBRANDT.Rda
Baseline Group	Rem_olig_grade2
Groups	Rem_olig_grade3

List Name:

Save Selected ↓

View HeatMap for selected reporters

These are the results of the group comparison.

You can sort this table based on any of the columns.

In this example, they are sorted based on fold change

Analysis Results									
<input type="checkbox"/>	Reporter ID	Gene Symbol	p-value	Fold Change ↕	Mean Baseline (log)	Mean Group (log2)	Std Baseline	Std Group	Target Data
<input type="checkbox"/>	209937_at	TM4SF4	$1.073 \times 10^{-3}$	12.039	5.403	8.993	2.563	1.988	
<input type="checkbox"/>	1553467_at	FLJ32742	$1.295 \times 10^{-3}$	9.743	6.182	9.466	2.384	1.901	
<input type="checkbox"/>	1557053_s_at	UBE2G2	$1.371 \times 10^{-3}$	7.776	5.409	8.368	1.787	2.266	
<input type="checkbox"/>	201981_at	PAPPA	$1.776 \times 10^{-3}$	7.771	7.376	10.334	2.265	1.779	
<input type="checkbox"/>	1555409_a_at	BAGE2	$1.378 \times 10^{-2}$	7.102	8.637	11.465	2.428	2.642	
<input type="checkbox"/>	208672_s_at	SFRS3	$9.744 \times 10^{-4}$	6.746	5.102	7.856	1.918	1.573	

# Going from Group Comparison to Gene expression KM

Analysis Results	
Statistical Method	TTest
Adjustment	FDR
Fold Change	1.5
Pvalue	.05
Study	REMBRANDT
Data File	REMBRANDT.Rda
Baseline Group	Rem_olig_grade2
Groups	Rem_olig_grade3

List Name:

Analysis Results	
<input type="checkbox"/>	Reporter ID
<input type="checkbox"/>	209937_at
<input type="checkbox"/>	1553467_at
<input type="checkbox"/>	1557053_s_at
<input type="checkbox"/>	201981_at
<input type="checkbox"/>	1555409_a_at
<input type="checkbox"/>	208672_s_at
<input type="checkbox"/>	1554593_s_at
<input type="checkbox"/>	208765_s_at
<input type="checkbox"/>	1555238_at
<input type="checkbox"/>	209997_x_at
<input type="checkbox"/>	215234_at

- Perform Gene Expression KM ▶
- Perform Gene Expression Search
- Search in Entrez
- View in UCSC Genome Browser
- Search in iHOP
- Search in PIR
- Search in Ensembl Gene View
- Search in Reactome
- View at KEGG
- View at QuickGO
- View at GeneCards
- View at String DB

Endpoint: Overall survival in months

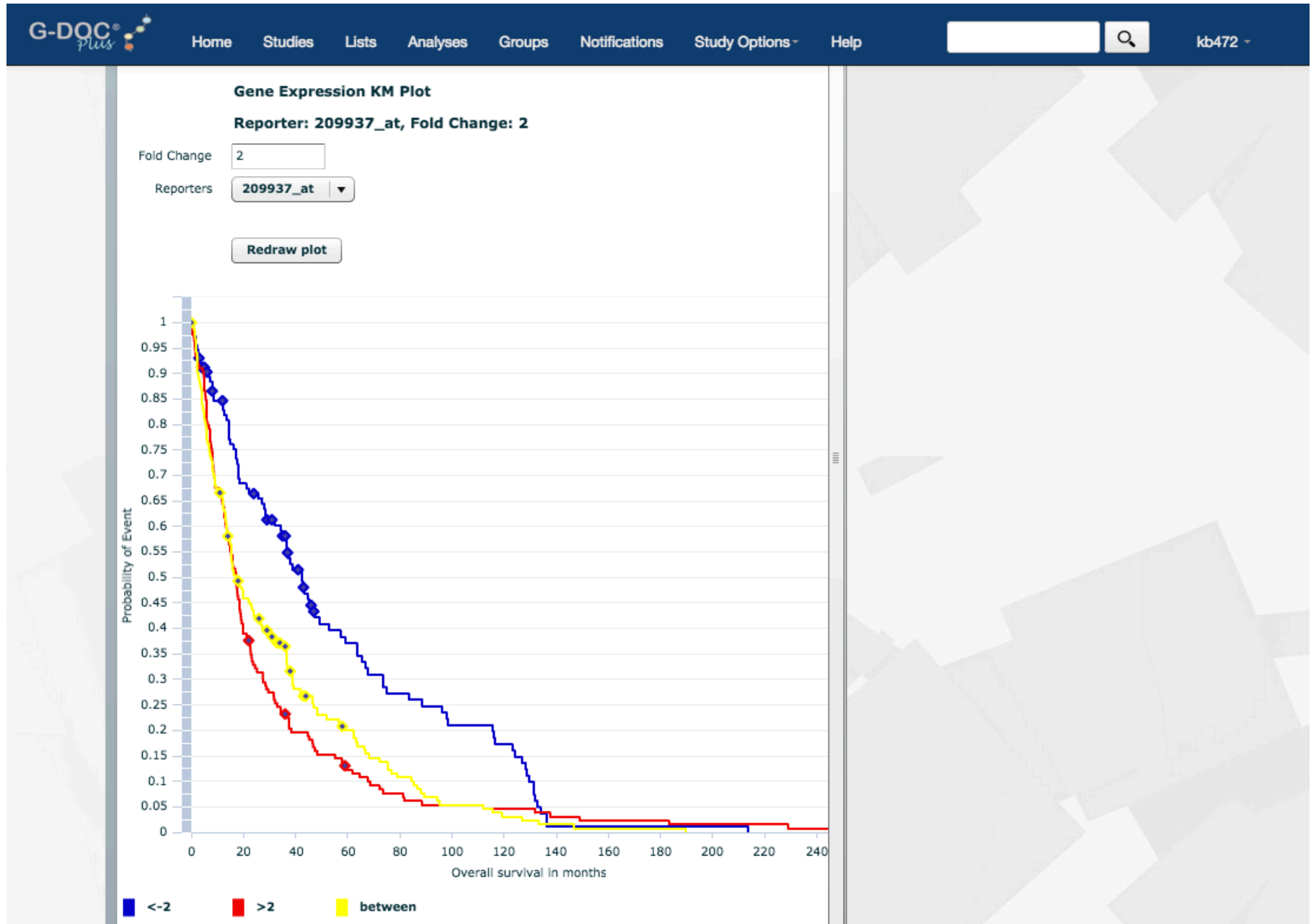
Fold Change	Mean Baseline (log	Mean Group (log2	Std Baseline	Std Group	Target Data
2.039	5.403	8.993	2.563	1.988	
743	6.182	9.466	2.384	1.901	
776	5.409	8.368	1.787	2.266	
771	7.376	10.334	2.265	1.779	
102	8.637	11.465	2.428	2.642	
746	5.102	7.856	1.918	1.573	
469	6.001	8.695	1.764	1.884	
434	5.398	8.084	1.843	1.989	
404	5.562	8.241	1.629	1.810	
328	6.215	8.877	1.709	1.521	
1.011 × 10 <sup>-2</sup>	5.660	8.205	1.137	2.165	

Going from Group Comparison to Gene expression KM:

- Click on gene of interest – Select “Perform Gene expression KM” – Select endpoint.

G-DOC allows users to perform Gene expression KM on either all samples or a group of samples

# Gene expression KM plot



Create group of interest, perform gene expression KM

## **METHOD 2**

# Steps


- Login
- Select study of interest (follow steps in slides 4–8)
- Create group of interest (follow steps in slides 9-12)
- Go to Study Options – Select “Gene expression KM plot”
- Select settings based on interest, click “plot”

See screenshots on next slide


G-DOC<sup>plus</sup> Home Studies Lists Analyses Groups Notifications Study Options Help kb472

# G-DOC Plus Launch Pad


Welcome! The G-DOC Plus Launch Pad is your one-stop resource for getting started on the platform.




Studies




Lists




Analyses



Groups



Notifications



Help me pick a study

Study Selected  
REMBRANDT

SEARCH  
Genome Browser  
Compound/Drug Targets  
Findings  
Explore Clinical Data and Create Groups  
Gene Expression Data

ANALYZE  
Group Comparison  
Chromosomal Instability Index  
KM Clinical Plot  
**KM Gene Expression Plots**  
Classification  
HeatMap Viewer



G-DOC<sup>plus</sup> Home Studies Lists Analyses Groups Notifications Study Options Help

# Create Gene Expression KM Plot

Current Study: REMBRANDT [change study?](#)

Select Patient Groups: Rem\_oligo\_gr2

Select Gene: PAPPA

Select Endpoint: Overall survival in months

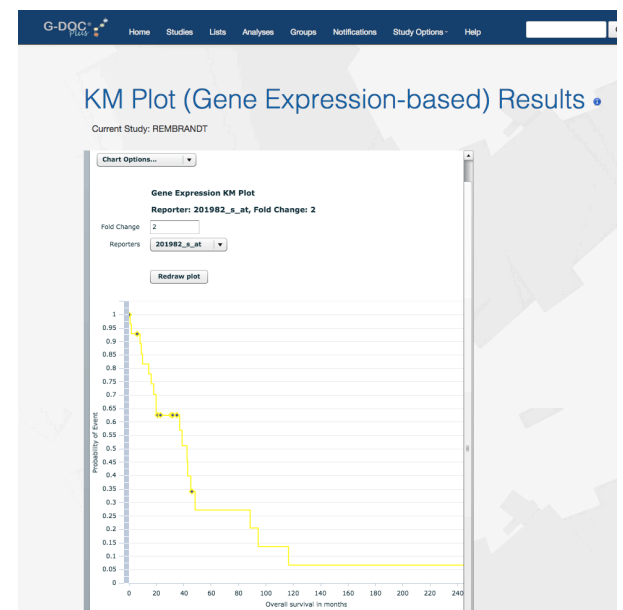
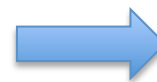
Data-Type: GENE EXPRESSION

Dataset: mas5 normalization

Plot

Reset

...performing gene expression analysis



Perform gene expression KM plot on all samples

## **METHOD 3**

# Steps

- Login
- Select study of interest (follow steps in slides 4–8)
- Go to “Study Options” – Select “Gene expression KM plot”
- Select settings based on interest, click “plot”

See screenshots on next slide



G-DOC Plus Launch Pad

Welcome! The G-DOC Plus Launch Pad is your one-stop re started on the platform.

Studies

Lists

Analyses

Groups

Notifications

Help me pick a study

Study Selected  
REMBRANDT

SEARCH

Genome Browser  
Compound/Drug Targets  
Findings  
Explore Clinical Data and Create Groups  
Gene Expression Data

ANALYZE

Group Comparison  
Chromosomal Instability Index  
KM Clinical Plot  
**KM Gene Expression Plots**  
Classification  
HeatMap Viewer



G-DOC Plus

Home Studies Lists Analyses Groups Notifications Study Options Help

## Create Gene Expression KM Plot

Current Study: REMBRANDT [change study?](#)

Select Patient Groups: All Patients

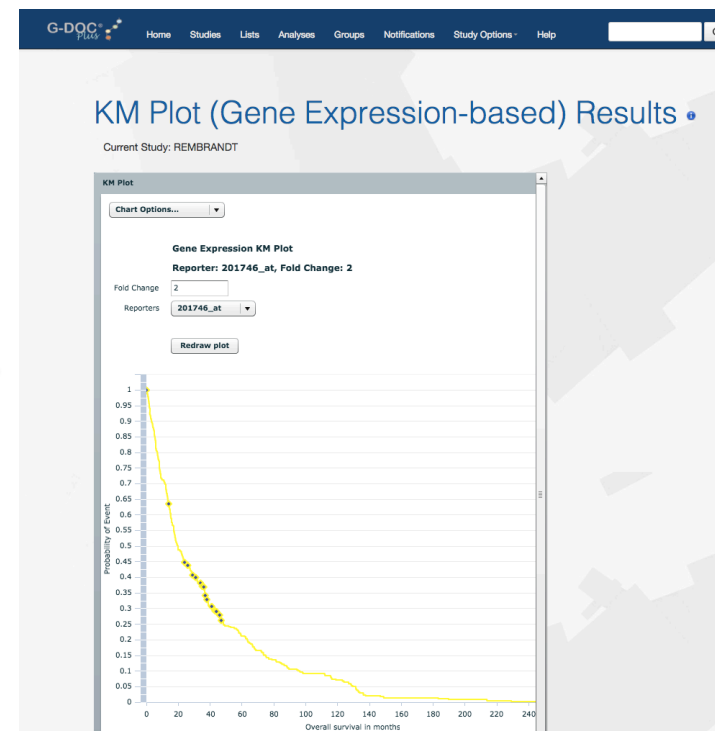
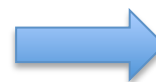
Select Gene: TP53

Select Endpoint: Overall survival in months

Data-Type: GENE EXPRESSION

Dataset: mas5 normalization

Plot Reset



# Changing settings for your KM plot

- No matter which method you use to create your KM plots, you can
  - Change “fold change” values and “Redraw” your plots to see how the patient stratification changes
  - Choose another “reporter id” (gene expression probe id) and “Redraw” your plots to see how the patient stratification changes
  - Save your plot – click on “Chart options” and you should

# Algorithm used in gene expression KM plot tool

- There are various ways of defining "low", "intermediate", and "high" expression. Some methods include: using mean or median as cut off; using 25% and 75% quartiles, etc.
- In G-DOC, we use "fold change" concept to split the gene expression data into "low", "intermediate" and "high expression".
  - The fold change is calculated for each sample in the input based on mean expression of a gene (indicating how many fold higher or lower the expression is compared to mean).
  - G-DOC lets users to choose the fold change cut off values. So if the cut off is 2, then samples with fold change  $< -2$  are equivalent to "low expression" (down-regulated). Samples with fold change between  $-2$  and  $2$  are equivalent to "intermediate expression". Samples with fold change  $> 2$  are equivalent to "high expression" (up-regulated).
- Note that the gene expression values are log2 values, so fold change 1 means ratio of 2.

**TIPS**

# General tips

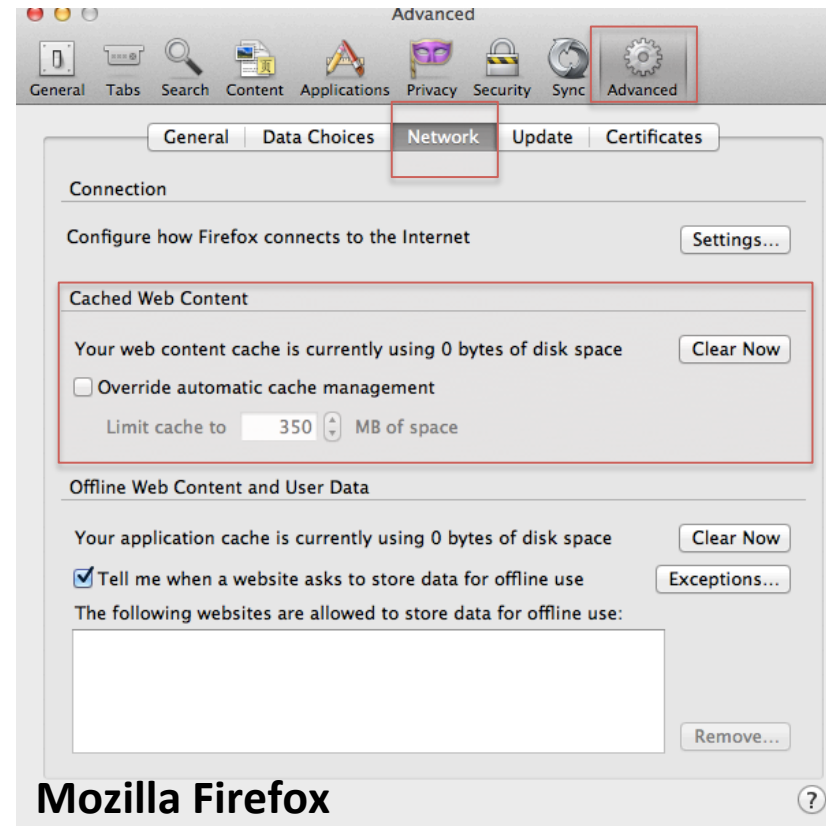
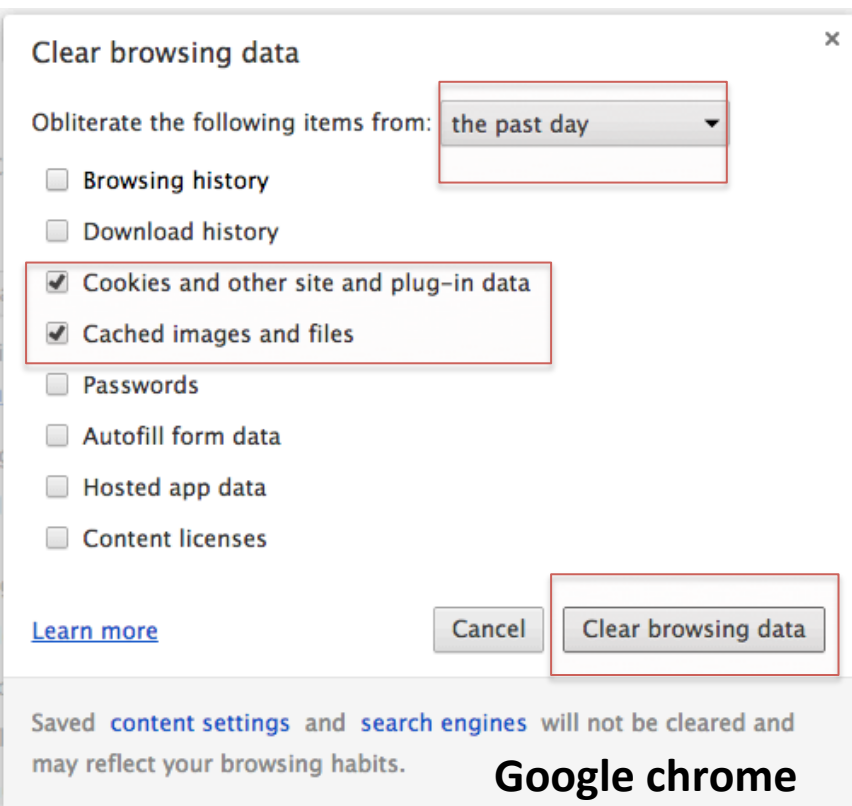
- G-DOC *Plus* works best if you don't use the **back** button in the web browser repeatedly.

Once you select a study, most tools will be easily available from the the top menu bar inside G-DOC *Plus*.

- The Pathway enrichment and the Lists tool may sometimes take a few seconds longer to execute than other tools (since they are directly connecting to the server every time). Your patience is highly appreciated.

# Clearing cache

- If the G-DOC web page does not respond after several seconds, try:
  - refreshing the page.
  - Log out and log back in, and try again
  - If the above two do not work, its possible that your web browser cache may need to be cleared
    - For Google chrome, go to **Settings** -> **Show Advanced Settings** -> Under “Privacy”, select **Clear Browsing data**
    - For Mozilla Firefox, go to **Preferences** -> **Advanced** -> **Network** -> Under “Cached Web Content” -> **Clear now**



- We are working hard to improve G-DOC *Plus*. Please feel free to email your questions and comments (no homework questions please) to us at: [gdoc-help@georgetown.edu](mailto:gdoc-help@georgetown.edu)